Selecting a Two-Group Classification Weighting Algorithm: Take Two

John D. Morris
Florida Atlantic University

Mary G. Lieberman
Florida Atlantic University

The two-group cross-validation classification accuracies of six algorithms (i.e., least squares, ridge regression, principal components, a common factor method, equal weighting, and logistic regression) were compared as a function of degree of validity concentration, group separation, and number of subjects. Therein, the findings of two previous studies were extended to the latter three methods with particular interest in how logistic regression faired as a function of validity concentration. In respect to validity concentration, as well as group separation and N, logistic regression was a mirror image of least squares. The same relative decrease, in respect to alternate methods, in accuracy with increasing validity concentration previously evidenced with least squares was observed. However, the large number of samples in which logistic regression failed to yield a solution may be a cause for concern.

This investigation extends Morris and Huberty (1987) by contrasting the accuracy of six algorithms for classifying subjects into one of two groups. Ordinary Least Squares (OLS), Ridge Regression (Ridge), Principle Components (PC), Pruzek and Frederick's (1978) common factor method (Pruzek), Equal Weighting (Equal), and Logistic Regression (LR) are the techniques that were compared.

Darlington (1978) posited that regression cross-validation accuracy is a function of R^2, N, and validity concentration, where R^2 represents the squared multiple correlation and N is the sample size. In Darlington's formulation, validity concentration was used to describe a data condition in which the principal components of the predictors with large eigenvalues also have large correlations with the criterion. Thus, validity concentration requires at least a modicum of predictor variable collinearity (i.e., large predictor eigenvalues), but collinearity is only necessary, not sufficient, for validity concentration. Darlington suggested that the most useful statistical technique for practical prediction problems, therein, may be ridge regression.

Through simulation, Morris (1982) re-examined the performance of Ridge with the same data structures on which Darlington posited the technique's superiority. With large validity concentration, Ridge was, indeed, more accurate than OLS, but contrary to Darlington's suggestions, Ridge was never the most accurate prediction technique. That is, in each case in which Ridge surpassed OLS due to large validity concentration, other contending weighting methods surpassed it.

Morris and Huberty (1987) examined a subset of the methods (e.g., OLS, Ridge, and PC) considered in Darlington (1978) and Morris (1982) in the context of two-group classification accuracy, rather than regression, using the same simulated data conditions, but extending them to three different population model accuracies and two different sample sizes; not including the Pruzek nor Equal methods as in Morris (1982).

The present study uses the same data conditions as in Morris and Huberty (1987), but includes the Pruzek and Equal methods. In addition, Logistic Regression (LR), an increasingly popular classification technique, was included. The same difficulties with multicollinearity (well-known with the OLS algorithm), have been mentioned (Hosmer & Lemeshow, 2000) as theoretically problematic to LR. Furthermore, ridge techniques have been suggested as a solution (Schaefer, 1986). However, the degree of effect of multicollinearity, and potentially resulting validity concentration, on the classification accuracy of LR has not been empirically and systematically examined.

Method

For easy comparison, all prediction algorithms that were included in the Morris and Huberty (1987) comparison (e.g., OLS, Ridge, and PC) were identically executed to those used in that comparison, and the Pruzek and Equal methods were identically calculated to that in Morris (1982). Specifically:

- **OLS**: Fisher's LDF was used.
- **Ridge**: Empirical ridge regression weights were calculated using the biasing parameter, *k*, due to Lawless and Wang (1976).
- **PC**: The principal components predictions were obtained with a regression model using all principal components with eigenvalues larger than one (i.e., Kaiser's rule).

Equal: A predictor was assigned a standardized weight of one and the same sign as the predictor-criterion correlation if the value of that correlation was statistically significant ($p < .01$). If no predictor-criterion correlation was this large, in absolute value, the predictor with the largest correlation was assigned a 1 with appropriate sign. This last step was executed so that there would be no case in which all predictors were assigned zero weights; thus, resulting in predicted values with no variance.

Pruzek: The common factor method of Pruzek and Frederick (1978) was used with rank determined by Kaiser's rule, as with PC, and communalities were iterated from squared multiple correlations.

LR: The logistic regression maximum likelihood solution was calculated with iteration using the Newton-Raphson method. Iteration was halted when all weights changed by no more than .001. In this subroutine, failure in solution generation was captured as caused by 1) complete separation, 2) a singular weight covariance estimate occurrence at an iteration, or 3) non-convergence of weights that were increasing to infinity.

The same fundamental data configurations that were originally posited by Darlington (1978) were used. Thus, all data conditions were such that there were 10 predictor variables. Four levels of collinearity were created so that the eigenvalues of the predictor variable intercorrelation matrix decreased by .50, .65, .80, and .95. Since the eigenvalues had to sum to 10, they were uniquely determined in each condition. Six levels of validity concentration were created and the component validities were proportional to a power of the eigenvalues, with powers of .1, .5, 1, 2, 4, and 10 considered. Darlington specified one population squared multiple correlation of .25. Thus, in this case, as the squared component validities had to sum to .25, they were also uniquely determined. Darlington also used one sample size of 40.

Procedure

As in Morris and Huberty (1987), these 24 conditions were expanded in a fully crossed design to include two more population multiple correlations of .75 and .95. As well, a sample size of 100 was also included and a total of 144 data conditions were considered. A population of 10,000 subjects, as opposed to 1,000 in Morris and Huberty, was created that manifested each of the 72 (4 X 6 X 3) desired sets of collinearity, validity concentration, and multiple correlation characteristics. Manipulating sample size was done by simply selecting 40, or alternatively 100, subjects in a sample with replacement.

The algorithm used to create the population covariance matrices representing the respective eigenstructure and validity concentration is described in Morris (1982), and the algorithm for creating the population of 10,000 subjects manifesting each of these desired covariance matrices is described in Morris (1975). As in Morris and Huberty (1987), to translate each of these data structures to a two-group classification problem, each population was dichotomized at the median criterion score and the prediction weights from randomly selected samples of the desired size were calculated to predict the dichotomous criterion by each of the six methods. The sample weights were then cross-validated by using them to classify all 10,000 population subjects. The total number of correct classifications was used to compare the relative accuracies of the methods. This procedure was replicated 100 times within each data condition with the mean classification accuracy representing the accuracy of each method. A Fortran 90 computer program compiled by Intel Parallel Studio XE 2011 was used for accomplishing all simulations.

Results

Tables 1, 2 and 3 show the results of all 144 simulation combinations. An orientation regarding validity concentration in the tables may be helpful. Within a table (i.e., thus, the population multiple correlation), as eigenvalue ratio becomes larger, multicollinearity becomes smaller. Within eigenvalue ratio, validity concentration becomes larger as power becomes larger. For each table, for example, the .50 eigenvalue ratio and .1 power offer the highest multicollinearity and lowest validity concentration considered.

Although statistical significance was not considered nearly as important as establishing trends over the data conditions, an overall Hotelling T^2 test and post hoc comparisons were used to contrast the methods' classification hit-rates as found in Morris and Huberty (1987). Thus, in line with Morrison's suggestion (1976), the means were ordered and pairwise post hoc contrasts were used to contrast larger means with smaller means until a statistically significant ($p < .01$) difference was found; thus, delineating best methods.

Table 1. Mean Proportion Correctly Classified for $R^2 = .25$.

Eigenvalue		Method											
Ratio	Power	N = 40						N = 100					
		OLS	Ridge	PC	Pruzek	Equal	LR/#*	OLS	Ridge	PC	Pruzek	Equal	LR/#*
.50	.1	**.5878**	.5699	.5480	.5455	.5460	**.5882**/100	**.6288**	.6049	.5735	.5704	.5803	**.6287**/100
	.5	.5868	**.6011**	.5899	.5856	.5917	.5871/100	**.6266**	**.6317**	.6152	.6116	.6228	**.6264**/100
	1.	.5875	**.6158**	**.6141**	**.6088**	**.6196**	.5867/99	.6262	**.6422**	.6390	.6352	**.6402**	.6258/100
	2.	.5904	**.6276**	**.6299**	.6248	**.6313**	.5895/99	.6294	.6497	**.6533**	.6506	**.6532**	.6285/100
	4.	.5911	.6286	**.6321**	.6280	**.6308**	.5893/99	.6261	.6488	**.6530**	.6509	**.6550**	.6258/100
	10.	.5909	.6279	**.6333**	**.6286**	**.6340**	.5892/98	.6271	.6483	.6519	.6504	**.6562**	.6266/100
.65	.1	**.5990**	.5853	.5579	.5594	.5636	**.5974**/100	**.6356**	.6229	.5836	.5808	.5947	**.6350**/100
	.5	**.5952**	**.5976**	.5818	.5810	.5843	**.5944**/100	**.6330**	**.6311**	.6078	.6054	.6176	**.6327**/100
	1.	.5930	**.6098**	**.6050**	**.6023**	**.6009**	.5929/98	.6316	**.6383**	.6279	.6249	**.6335**	.6314/100
	2.	.5914	**.6169**	**.6227**	.6165	.6135	.5906/97	.6271	**.6404**	**.6438**	**.6398**	**.6452**	.6268/100
	4.	.5914	.6211	**.6295**	.6231	.6182	.5905/99	.6286	.6437	**.6504**	.6478	**.6496**	.6282/100
	10.	.5920	.6219	**.6309**	.6250	.6200	.5912/99	.6273	.6437	**.6499**	.6480	.6470	.6267/100
.80	.1	**.5898**	**.5836**	.5605	.5748	.5558	.5900/99	**.6283**	.6220	.5817	.5908	.5868	**.6279**/100
	.5	**.5894**	**.5888**	.5707	**.5820**	.5611	.5894/99	**.6276**	.6244	.5928	.5989	.5975	**.6272**/100
	1.	**.5898**	**.5958**	.5851	**.5917**	.5702	.5899/99	**.6282**	**.6283**	.6106	.6134	.6121	**.6277**/100
	2.	.5898	**.6028**	**.6025**	**.6032**	.5827	.5905/99	.6265	**.6313**	**.6285**	.6269	.6251	.6260/100
	4.	.5898	**.6057**	**.6134**	**.6081**	.5850	.5903/100	.6253	.6331	**.6400**	.6368	.6303	.6248/100
	10.	.5908	.6100	**.6231**	.6136	.5884	.5906/99	.6265	.6357	**.6449**	.6411	.6289	.6261/100
.95	.1	**.5871**	**.5843**	.5607	**.5849**	.5634	.5858/98	**.6252**	.6223	.5855	**.6223**	.5941	**.6253**/100
	.5	**.5874**	**.5842**	.5623	**.5848**	.5642	.5862/98	**.6262**	.6237	.5883	.6230	.5940	**.6263**/100
	1.	**.5876**	**.5843**	.5624	**.5849**	.5637	.5866/99	**.6261**	**.6241**	.5910	**.6232**	.5942	**.6261**/100
	2.	**.5876**	**.5855**	.5652	**.5868**	.5657	.5867/99	**.6261**	**.6244**	.5951	**.6231**	.5954	**.6262**/100
	4.	**.5872**	**.5867**	.5688	**.5873**	.5661	.5871/99	**.6258**	**.6240**	.6006	**.6224**	.5978	**.6259**/100
	10.	**.5850**	**.5881**	**.5760**	**.5874**	.5642	**.5848**/100	**.6259**	**.6259**	.6139	**.6260**	.6031	**.6255**/100

Note: The best performing method or set of methods is in bold ($p<.01$).

* Number of samples for which LR calculable; not included in contrast if any not calculable.

Table 2. Mean Proportion Correctly Classified for R^2 = .75.

Eigenvalue		Method											
Ratio	Power	N = 40						N = 100					
		OLS	Ridge	PC	Pruzek	Equal	LR/#*	OLS	Ridge	PC	Pruzek	Equal	LR/#*
.50	.1	**.7560**	.7064	.6248	.6139	.6375	.7449/45	**.7982**	.7564	.6591	.6495	.6649	**.8003**/100
	.5	**.7595**	**.7681**	.7197	.7080	.7326	.7514/50	**.8042**	.8011	.7446	.7410	.7247	**.8062**/100
	1.	.7615	**.7948**	.7734	.7628	.7721	.7487/57	.8054	**.8162**	.7947	.7905	.7580	.8069/100
	2.	.7634	**.8031**	**.8066**	.7977	.7950	.7466/49	.8046	**.8184**	**.8205**	.8162	.7935	.8051/100
	4.	.7626	.8035	.8133	.8057	**.8143**	.7532/47	.8057	.8201	**.8270**	.8241	.8227	.8066/100
	10.	.7609	.8010	.8115	.8045	**.8154**	.7487/53	.8027	.8160	.8219	.8196	**.8250**	.8039/100
.65	.1	**.7613**	.7391	.6410	.6456	.6660	.7616/41	**.8049**	.7916	.6647	.6709	.6922	**.8061**/100
	.5	**.7619**	**.7606**	.6916	.6943	.7161	.7572/50	**.8053**	**.8031**	.7183	.7261	.7296	.8057/98
	1.	.7619	**.7771**	.7425	.7418	.7568	.7473/50	.8047	**.8092**	.7662	.7693	.7638	.8053/100
	2.	.7715	**.7977**	**.7938**	.7871	**.7918**	.7622/51	.8067	**.8155**	.8110	.8079	.7905	.8081/100
	4.	.7677	.7994	**.8106**	.8003	.8004	.7512/53	.8087	.8174	**.8264**	.8227	.8119	.8092/100
	10.	.7677	.7994	**.8123**	.8034	**.8065**	.7556/51	.8066	.8150	**.8244**	.8218	**.8260**	.8071/100
.80	.1	**.7585**	**.7506**	.6482	.7016	.6760	.7502/58	**.8015**	.7984	.6748	.7320	.7077	**.8024**/100
	.5	**.7598**	**.7574**	.6744	.7152	.6947	.7485/59	**.8042**	**.8034**	.7034	.7504	.7255	**.8048**/100
	1.	**.7565**	**.7605**	.7046	.7295	.7155	.7494/64	**.8047**	**.8056**	.7352	.7680	.7469	**.8047**/100
	2.	.7582	**.7709**	.7486	.7572	.7533	.7581/57	.8048	**.8075**	.7772	.7918	.7819	.8043/100
	4.	.7527	**.7738**	**.7840**	**.7780**	.7616	.7424/51	.8045	**.8084**	**.8091**	**.8107**	.7942	.8045/100
	10.	.7627	.7819	**.7971**	.7847	.7585	.7489/46	.8024	.8064	**.8170**	.8121	.8119	.8025/100
.95	.1	**.7559**	**.7541**	.6563	.7389	.6868	.7442/44	**.7978**	**.7973**	.6877	.7913	.7623	**.8001**/100
	.5	**.7551**	**.7533**	.6575	.7389	.6831	.7391/45	**.7981**	**.7978**	.6910	.7923	.7603	**.8005**/100
	1.	**.7560**	**.7549**	.6610	.7421	.6881	.7446/44	**.7984**	**.7980**	.6957	.7923	.7604	**.8011**/100
	2.	**.7526**	**.7529**	.6647	.7399	.6876	.7485/43	**.7977**	**.7976**	.7028	.7921	.7570	**.8006**/100
	4.	**.7511**	**.7531**	.6730	.7399	.6834	.7429/53	**.7950**	**.7951**	.7167	.7904	.7551	**.7969**/100
	10.	**.7538**	**.7576**	.6942	.7457	.6966	.7434/50	**.8017**	**.8023**	.7505	.7989	.7749	**.8039**/100

Note: The best performing method or set of methods is in bold ($p<.01$).

* Number of samples for which LR calculable; not included in contrast if any not calculable.

Table 3. Mean Proportion Correctly Classified for $R^2 = .95$.

Eigenvalue Ratio	Power	N = 40 OLS	Ridge	PC	Pruzek	Equal	LR/#*	N = 100 OLS	Ridge	PC	Pruzek	Equal	LR/#*
.50	.1	**.8415**	.7747	.6556	.6263	.6774	.8649/4	**.8873**	.8312	.6928	.6808	.6905	.9009/58
	.5	**.8407**	**.8426**	.7637	.7472	.7731	.8271/2	**.8885**	.8812	.7896	.7866	.7541	.8982/66
	1.	.8480	**.8756**	.8377	.8276	.8165	.8132/2	.8922	**.9041**	.8608	.8565	.7962	.9028/54
	2.	.8429	**.8848**	**.8881**	.8745	.8626	.8316/1	.8919	**.9071**	**.9068**	.8992	.8535	.8987/53
	4.	.8431	.8851	**.8979**	.8889	**.8986**	.8737/2	.8885	.9035	**.9117**	.9067	.9014	.8976/67
	10.	.8450	.8850	.8997	.8884	**.9069**	.8671/4	.8903	.9051	.9133	.9085	**.9182**	.8990/61
.65	.1	**.8441**	.8117	.6688	.6750	.7060	.8830/3	**.8901**	.8769	.6938	.7027	.7213	.9007/57
	.5	**.8460**	**.8428**	.7357	.7396	.7659	.8866/1	**.8913**	**.8889**	.7590	.7723	.7625	.8999/59
	1.	.8444	**.8589**	.8007	.8011	.8092	.7995/1	.8917	**.8967**	.8238	.8334	.8068	.9016/62
	2.	.8493	**.8769**	**.8705**	.8633	.8533	.8281/1	.8951	**.9032**	**.8917**	.8896	.8465	.9039/60
	4.	.8489	.8783	**.8961**	.8816	.8763	.8891/1	.8934	.9021	**.9149**	.9080	.8791	.9044/54
	10.	.8433	.8763	**.8973**	.8840	**.8912**	.8437/5	.8904	.8989	**.9124**	.9064	**.9122**	.8993/60
.80	.1	**.8434**	**.8362**	.6765	.7563	.7217	.8541/4	**.8920**	.8892	.7070	.7991	.7470	.9010/55
	.5	**.8431**	**.8406**	.7103	.7774	.7419	.0000/0	**.8896**	**.8889**	.7408	.8164	.7686	.9008/54
	1.	**.8415**	**.8459**	.7525	.7981	.7753	.8460/3	**.8903**	**.8908**	.7824	.8382	.7978	.8997/65
	2.	.8421	**.8533**	.8126	.8325	.8299	.8584/4	.8932	**.8949**	.8428	.8709	.8441	.9021/52
	4.	.8389	**.8558**	**.8577**	**.8568**	.8410	.0000/0	.8928	**.8961**	**.8922**	**.8972**	.8557	.9016/52
	10.	.8407	.8605	**.8814**	.8650	.8351	.8555/2	.8940	.8973	**.9093**	.9021	.9021	.9030/49
.95	.1	**.8391**	**.8383**	.6919	.8128	.7568	.8608/2	**.8860**	**.8860**	.7237	.8707	.8225	.8976/57
	.5	**.8384**	**.8383**	.6946	.8129	.7533	.0000/0	**.8859**	**.8859**	.7277	.8708	.8230	.8986/57
	1.	**.8380**	**.8382**	.6974	.8134	.7528	.7810/1	**.8869**	**.8868**	.7328	.8713	.8242	.8982/57
	2.	**.8380**	**.8386**	.7033	.8140	.7501	.8189/3	**.8863**	**.8864**	.7438	.8720	.8222	.8996/62
	4.	**.8329**	**.8351**	.7126	.8107	.7469	.8176/3	**.8837**	**.8840**	.7624	.8703	.8205	.8965/64
	10.	**.8354**	**.8395**	.7388	.8182	.7762	.8306/6	**.8873**	**.8879**	.8044	.8772	.8327	.8991/61

Note: The best performing method or set of methods is in bold (p<.01).

* Number of samples for which LR calculable; not included in contrast if any not calculable.

The obvious effect of increased accuracy as the population multiple correlation or sample size increased is evident. As would be expected, the trend regarding validity concentration was parallel to that found in Morris and Huberty (1987) in respect to OLS, Ridge, and PC. OLS was superior at smaller levels of validity concentration. As validity concentration increased within an eigenvalue ratio level, Ridge became superior, and if the validity concentration became large enough (note that the eigenvalue ratio limited the validity concentration), PC eventually became superior. These results are consistent with prior literature (with a different Fortran compiler, word length, random seed, and random number generator) and should provide increased comfort regarding the performance of the methods previously examined.

In respect to Pruzek and Equal, there is also similarity with regression results (Morris, 1982). With increasing validity concentration, Pruzek appears to function much as does PC, exceeding the performance of OLS and LR, and eventually Equal becomes superior at the very largest validity concentration levels. In respect to OLS, this finding is consistent, although perhaps less dramatic than in Morris (1982) in the regression case. This may be due to the lessened variance of the dichotomous criterion. In addition, the fact that Pruzek's common factor method is using dimensionality delivered by the Kaiser eigenvalue-less-than-one dimensionality rule, derived from a principal components solution, may disadvantage the technique. Other research shows that prediction from a common factor perspective (Morris & Guertin, 1977), and specifically from extensions of the Pruzek method (Pruzek & Lepak, 1992), are particularly meritorious.

It seems that our canned admonitions regarding the hazards of multicollinearity in the use of OLS methods may need a bit of refinement. For example, an examination of results for the .50 eigenvalue ratio and .1 power (again, the largest degree of multicollinearity, but smallest degree of validity concentration) in each table illustrates that OLS, along with LR, performs just fine and, indeed, is always superior to other methods. This result is consistent with the two-group classification results in Morris and Huberty (1987) and with the regression results in Morris (1982). One can see that within an eigenvalue ratio, other methods only become superior to OLS and LR as power, and thus validity concentration, becomes larger. At the largest eigenvalue ratio, this occurs quickly with increasing validity concentration. That is, as the eigenvalue ratio becomes larger, thus representing lower levels of multicollinearity, the process is slowed, or at the larger population multiple correlations, stopped. Multicollinearity is a necessary, but not sufficient, condition for validity concentration; yet, it is high validity concentration, not multicollinearity, which is problematic for OLS herein.

LR deserves special attention. In situations in which there were no failures to derive maximum likelihood weights, LR mirrored OLS essentially perfectly. Across all such data conditions, there was no case in which LR was significantly different than OLS. Though, one might surmise that, as LR does not require the same distributional assumptions as OLS, it may fair even better with non-normal data (Joachimsthaler & Stam, 2007; Press & Wilson, 1978). Although, in the case of multivariate normal data,

OLS is more efficient (Efron, 1975). Yet, these comparisons are anecdotal with specific data sets rather than a general examination accomplished by simulation varying population characteristics. This, then, means that all of the results regarding the effects of multicollinearity and validity concentration on OLS classification accuracy are also identically true for LR. The alternative methods that improve on OLS accuracy with higher collinearity and validity concentration also improve on LR to essentially the same degree and in the same pattern. As well, it is the effect of validity concentration that affects the performance of LR and not multicollinearity without validity concentration.

However, as can be seen, there were many failures of the LR method to derive a model. It is well known that in cases in which there is complete separation between groups, maximum likelihood weights are not possible; thus, the logistic regression algorithm will necessarily fail to resolve to a weight solution. There are alternatives, including exact logistic regression and penalized logistic regression (Firth, 1993; Heinze & Schemper, 2002). However, as interest in this study was in the performance of standard logistic regression routines available in standard statistical packages, such as SPSS; SAS; or Minitab, that typical researchers would elect to use, those techniques were not considered. It is also the case that in situations in which there is "near" complete separation, LR may fail to iterate to a model. This was the case herein in many simulation conditions. When LR failed to develop a model for a particular sample, the classification accuracy could, of course, not be included in the aggregate. Therefore, the number of samples for which LR models were able to be used to classify, thus aggregate, is included in the tables. Also, in those cases, LR was excluded from statistical significance testing as it manifested missing data and sometimes a lot of missing data.

For the smallest squared multiple correlation of .25 (i.e., representing the least group separation with the mean Mahalanobis D^2 = .79), results were not too bad, with only 1 to 3 samples for which LR weights were unavailable due to the algorithm's failure in the smaller sample size. However, as population group separation became larger, but by no means complete, failure increased dramatically. It was always worse with the smaller sample size, but certainly of importance at the larger sample size. At the .75 squared multiple correlation (Mean D^2 = 3.69), LR produced models for only about half of the samples for the N = 40 sample size across all eigenvalue ratio and validity concentration power conditions. For the N = 100 sample size, only in one instance was there LR regression model generation failure. However, at the .95 squared multiple correlation condition (D^2 = 6.22) for the N = 40 sample size, the largest number of models produced was 5 out of 100 and in two conditions, LR was unable to produce any models for the 100 samples. At this level of group separation, even at the N = 100 sample size, models were produced for no more than two-thirds of the samples.

Discussion

The performance of the additional methods, Pruzek and Equal, showed some similar performance characteristics to former studies. Pruzek and Equal improved on OLS classification accuracy with increasing validity concentration. In addition, whatever can be said of OLS can also be said of LR, as their performance was so similar, under conditions of high validity concentration, LR is surpassed by the alternate techniques sequenced in the same way as with OLS.

Further consideration regarding the technical specifics of the LR failures is necessary. A few of the failures were due to the estimated weight covariance matrix becoming singular. Most were due to iterations exceeding the programmed default of 20. However, it made no difference to increase that limit all the way to 100. The reason that the iteration limit was being exceeded was that the weights were increasing to infinity rather than converging. To allow the number of iterations to become essentially limitless, at 100, caused the failure to then occur due to a subsequent singularity of the weight covariance matrix. As is often the case in "runaway" maximum likelihood estimation, the weights are not converging because they, or some subset of them, are increasing to infinity. Although increasing the iteration limit made no difference, results using a limit of 20 are presented as that is the default limit for SPSS. Extensive checking with SPSS LR was done on the same sampled data sets with the same results.

A question of interest regards how do these data conditions fit with what the typical researcher encounters and how often do such failures accrue in practice? An additional difficulty is that LR programs (e.g., SPSS, SAS, and Minitab) that fail with a data set, still report the weight set from the last iteration; albeit, with a note that iteration failed and no solution is possible. To its credit, STATA does not report an erroneous vector of weights if maximum likelihood convergence is not possible. STATA is, however, not the mainstream software used for LR. Another concern is whether those erroneous results are then reported by the naïve researcher.

However, if one is careful and considers the footnote, it is obvious that LR failed to converge and therefore there is no LR model to summarize, nor LR weights, even though SPSS proffers such. The

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.000a	.726	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1a	gpa	67.625	44793.133	.000	1	.999	2.341E29
	rfcat	10.771	23052.901	.000	1	1.000	47625.932
	mfcat	7.485	11430.954	.000	1	.999	1781.776
	schoolsat	3.991	13653.445	.000	1	1.000	54.117
	sri	.070	130.760	.000	1	1.000	1.073
	retentions	32.383	34125.344	.000	1	.999	1.158E14
	Constant	-294.890	116757.684	.000	1	.998	.000

a. Variable(s) entered on step 1: gpa, rfcat, mfcat, schoolsat, sri, retentions.

Figure 1. An example of an LR output in SPSS.

maximum likelihood iterations failed to converge and these are merely the weights calculated in the 20th iteration. The same failure is obtained if the maximum number of iterations is specified as 100 or other larger numbers.

These results would support a recommendation that, for data with low validity concentration, OLS or LR, if appropriate attention is paid to model generation failure, are superior and relatively equal choices. Although, for anomalous data conditions, it may be that LR may have the edge. As validity concentration increases, Ridge and reduced rank (i.e., PC or Pruzek) become top choices. At the very highest validity concentration, the simple EW algorithm is the best choice. In addition, this trend is lessened as sample size and degree of group separation increases.

References

Darlington, R. B. (1978). Reduced variance regression. *Psychological Bulletin*, *85*, 1238-1255.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, *70*, 892-898.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*, 27-38.

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*, 2409-2419.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Joachimsthaler, E. A., & Stam, A. (2007). Four approaches to the classification problem in discriminant analysis: An experimental study. *Decision Sciences*, *19*, 322-333.

Lawless, J. F., & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics*, *A5*, 307-323.

Meshbane, A., & Morris, J. D. (1996, April). *Predictive discriminant analysis versus logistic regression in two-group classification problems*. Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Morris, J. D. (1975). A computer program to create a population with any desired centroid and covariance matrix. *Educational and Psychological Measurement*, *35*, 707-710.

Morris, J. D. (1982). Ridge regression and some alternative weighting techniques: A comment on Darlington. *Psychological Bulletin*, *91*, 203-210.

Morris, J. D., & Guertin, W. H. (1977). The superiority of factor scores as predictors. *Journal of Experimental Education*, *45*, 41-44.

Morris, J. D., & Huberty, C. J. (1987). Selecting a two-group classification weighting algorithm. *Multivariate Behavioral Research*, *22*, 211-232.

Morrison, D. F. (1976). *Multivariate statistical methods*. New York: McGraw-Hill.

Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, *73*, 699-705.

Pruzek, R. M., & Frederick, B. C. (1978). Weighting predictors in linear models: Alternatives to least squares and limitations of equal weights. *Psychological Bulletin*, *85*, 2, 254-266.

Pruzek, R. M., & Lepak, G. M. (1992). Weighted structural regression: A broad class of adaptive methods for improving linear prediction. *Multivariate Behavioral Research*, *27*, 95-129.

Schaefer, R. L. (1986). Alternate estimators in logistic regression when the data are collinear. *Journal of Statistical Computation and Simulation*, *25*, 75-91.

Send correspondence to: John D. Morris
Florida Atlantic University
Email: jdmorris@fau.edu